dysfunction.

Treatment is dramatically different in systolic and diastolic dysfunction. If they don't have symptoms, the test isn't a very good predictor of ejection fraction so it still doesn't tell you who to treat.

I was curious because the sponsor did provide a literature review and one of the papers in the literature review I think may accurately reflect what may be a way of thinking about this. It's in Volume II, page 524, conclusion to paper submitted by the sponsor.

The determination of natriuretic peptides does not further increase the correct detection of LV ejection fraction, but it will improve the correct prediction of normal LV function. If that is the case, this is really a test that diagnoses health, not disease.

DR. MAISEL: Let me comment because I now that paper pretty well. I partially agree with you. I think if you can get a test with a really good negative -- in our group if we would have just studied people who just got echos with people with BNPs of 37 or lower, we would have been fine and not having to screen them so I do agree with that.

I think, however, that is one facet and that is a great reason to have a test. It's a great reason in the emergency room, Dr. Packer, when you don't have an echocardiogram and they have symptoms of heart failure to be able to be pretty sure. You know, you can't get an echo down there most of the time. In that setting to assist the patient, it's going to have a good negative predictive and it's going to have a good positive.

I think after that, and I really agree with everything you said, it's going to be a matter of once we explore and once we have patients in there with the echo diagnosis of systolic or diastolic dysfunction and then we want to give treatment, how are we going to follow those patients? I suspect that in the future even though I don't believe they are asking for that approval today, that's probably what we are going to use.

DR. PACKER: The application is suggesting that the approval be based on a guide to the diagnosis of heart failure but that's not what the data supports, or the diagnosis or the treatment of heart failure. It's really a guide to the identification of people who someone might suspect has heart failure and the test can tell them that they do not have heart

failure.

DR. MAISEL: I think the positive predictive value in --

DR. PACKER: It's not based on an adequate control group.

DR. MAISEL: Maybe we can explore this a little bit. I think, you know, in our 250-patient study those are adequate controls and it helped. It not only said who didn't but it basically said who did despite what the emergency department said.

I think also that even in all of John's slides here, even when you saw changes between men and women, even when you saw changes between renal and dysfunction and no renal dysfunction, even when you saw changes between people young and old, you get the sensation that those ranges are all falling somewhere between about 40 and 80.

In fact, in our emergency population that was right where that cutoff was, at about 80. I think when you get above that -- and I do think more studies need to be done but, I mean, there is still a big problem in diagnosing heart failure in just seeing what happened in the emergency department at our hospital and what we see in the echo lab.

I think that we've pulled a bunch of

patients that had abnormal echos that had no symptoms. We just wanted to screen them and they had BNPs of 400 and they have low EFs or they have diastolic dysfunction.

I do think we need to separate out systolic from diastolic. BNP obviously can't do it but as a trigger to get in the system because both of those, now even with diastolic dysfunction, there's a pretty significant mortality and I think we have to get them in the system.

DR. PACKER: I don't disagree but this test doesn't save one echo from being done appropriately because you still need the echo to distinguish between systolic and diastolic dysfunction. You still need the echo to determine if there are valvular abnormalities or other structural disease that can be contributing to heart failure.

It isn't a test that reduces health care cost. If anything, it's a test that increases health care cost because it doesn't -- you know, you still have to do all of these tests because it doesn't replace those tests. It doesn't provide incremental information above and beyond what we need to derive from the test that we normally do.

DR. MAISEL: I don't want to speak for the

company. I don't think they are necessarily asking that this be approved as a screening test. I think they are asking this be approved for an aid to diagnosis. An aid to diagnosis, I think, just means that, can it help you diagnose it.

When you say can't something helps you diagnose it, it means either it helps you diagnose it as heart failure or it helps you diagnose it as something else. And diagnosing it as something else is just as important.

I mean, for the health care system if you miss a diagnosis in the emergency room, think about the health care expenditure there. Think about the hospitalization and possible morbidity and mortality from not having them on ACE inhibitors or beta blockers.

I think some of the things that were presented and some of the things that were talked about and things you are asking about are all things that I think really need to be explored. I think that judging from talking to people at the ACC, those are all sort of being explored right now.

Here right now you have something that is point-of-care which nothing else is out there. You have something you can stick right in an emergency

room and, again, I don't mean to get on a soapbox for this but you have something right down in the emergency room you could help patients.

You could help them by making a diagnosis or you could help them by ruling out. I think that is a very important thing. I think it does correlate loosely with ejection fraction. I think it correlates much more strongly with New York Heart Association Classification.

DR. PACKER: We don't need anything that correlates with New York Heart class because we can do that at the bedside.

DR. MAISEL: Well, you know --

DR. PACKER: New York Heart class is a clinical bedside evaluation.

DR. MAISEL: But, you know what? I'll tell you --

DR. PACKER: Just to make one more point because I really want to stop talking, but if one forces the workup of elderly patients with high BNP, one has to weigh that against missing the diagnosis in some and over diagnosing people in others.

The worse thing that could happen is that if the physician sees an elderly person with a 200 or 300 BNP level, treats that patient for heart failure

and that patient doesn't have heart failure, I'm not exactly certain what's worse.

DR. MAISEL: Hopefully they wouldn't just treat on the basis of one level, but it would get them in the system. Let me make just one other comment for New York Heart class. We just finished a study and we're writing up 72 patients but we have about 150 patients that were admitted for decompensated heart failure. We just didn't tell anybody the BNP level.

We just let them treat whatever they wanted and we're just looking at outcomes related to the BNP levels. You know, people go home and what happens? They get readmitted. Now, when do they go home? They go home when they feel better. That's reflected in the New York Heart classification.

What we have found and I think the best part of this study is that patients that did not come back to the hospital in the next 30 days or die at the hospital, their BNP levels all went down in the hospital, their New York Classification all improved in the hospital.

The patients that either died in the hospital or were readmitted within 30 days, the New York Heart class of people who at least went home who were readmitted, the New York Heart class all improved

in those patients because that's when you send them home. The BNP did not go down in those patients. The curve was flat. It just didn't budge.

I think we're not necessarily talking about -- we have a New York Heart class. We're talking about something that might help us modify treatment to those patients in the hospital and say, "I feel better." Get them out of there.

You know taking care of patients you're always fighting with your house staff to keep them in a couple extra days so you can get that ACE up. I think this is going to be a great way to prove that point that we want to get better treatment on board to keep them from coming back.

DR. KROLL: We need to give Dr. Comp an opportunity to ask a question.

DR. COMP: I'm familiar with the VA health care system. I'm a little concerned about the general applicability of people coming in with heart failure. Will this sort of data -- here you have people with peptide levels over 1,000 and they sound like they're Class IV. How is that going to help me with my little old ladies that have kind of pedal edema maybe from venue insufficiency, maybe from heart failure. They are a little short of breath.

DR. MAISEL: There is a panel of four slides. We looked at those little old ladies with edema because I think it's a big problem. It turns out, first of all, I think we're making too much of little old ladies here. I think the BNPs are up but they are not really up. I mean, they are somewhere between 40 and 80 but they're not 300, they're not 500, they're not 600, they're not 700.

There is a panel in there of patients with or without edema that we specifically looked at. Now, they didn't come in to get in the study. If they just came in with edema, "Doc, I've got this swelling," we couldn't put them in the study unless they had shortness of breath because that was the initial criteria. But if they did come in with edema and they had symptoms of shortness of breath and was in the study, the BNP levels were again ten-fold difference whether they had a final diagnosis of CHF or not.

DR. COMP: Just one other question. What happened to all the COPDers? I assume that 65 percent of the people don't have congestive heart failure. They have exacerbation of COPD.

DR. MAISEL: A lot of our patients that weren't congestive heart failure were, in fact -- that's the most common differentiation we saw. Most

of the patients -- 3/4 of the patients were either CHF or they were pulmonary disease with exacerbation. Some of those guys got admitted and some didn't. If we could get that panel up, the patients with lung disease -- I think this is a really telling point -- their BNPs aren't particularly high.

DR. COMP: What I'm saying is there weren't many COPDers in your study.

DR. MAISEL: Well, we had a lot of people when -- they got a primary classification. If they came in with pneumonia, they got the final diagnosis of pneumonia. A lot of those people had underlined COPD. Our total underlined population was almost all -- it was over 3/4 COPD.

DR. COMP: Not to belabor it but were their peptide levels ever measured? Did they get into your study is my point.

DR. MAISEL: Absolutely. Yes. The only ones that we excluded -- now, we didn't want to take people where it was clear they did not have even a chance of congestive heart failure. In other words, if we got a 32-year-old asthmatic that came in wheezing, we didn't put them in the study because we would be cheating favorably toward ourselves really because that guy was going to be normal and there is

no suspicion of heart failure.

We left it up to the ED physicians. We said these are the people we want and this is when we had our meeting for the multi-center trial. These were what the ED advisers wanted. they wanted people in which you could at least conceive of there being a possibility of CHF.

I'll tell you that going down in the ER in a VA population, that can be very hard to differentiate so I think a lot of those COPD guys get in there. They get in the study. Most of them. We checked the ICD codes for COPD at the end and 75 percent of everybody got in.

DR. KROLL: Dr. Comp, if you have additional questions, certainly when we come to the open committee discussion, we can ask any other questions of the sponsors and I think that's an important time to do that.

What I would like to do now is we need to get ready for a break. I want to make a comment first so we don't waste time on this issue. In the FDA, not considering cost in terms of this evaluation, because right now we're trying to consider safety and efficacy so we're not interested in cost concerns and whether it's going to save money in terms of saving other

diagnostic tests or things like that.

Let me turn it over to Veronica.

MS. CALVIN: Just one announcement -- two announcements actually. Anybody who has not signed in, please you need to do so at the registration desk. That includes panel members. Also, please help us keep the conference room clean. We have trash cans at the door. Thank you.

DR. KROLL: Now we'll take a 15 minute break after which there will be an FDA presentation.

(Whereupon, at 11:51 a.m. off the record until 12:12 p.m.)

DR. KROLL: This is Dr. Kroll. We'd like to get started now so the FDA can do their presentation so if everybody could please take their seats.

MS. CHESLER: Are we ready to start?

DR. KROLL: Yeah, why don't you go ahead and start.

MS. CHESLER: Okay. I'm going to be starting the FDA presentation. Good morning. I'm Ruth Chesler, Scientific Reviewer for the Chemistry Toxicology Branch and a member of the team reviewing this device. These are the members of the review team.

I'll be presenting a summary of the basic principles of this device and the studies that were conducted. Next, Dr. Marina Kondratovich, our statistician, will summarize observations from her review. Finally, I'll return to present the questions we would like you to address during your deliberations.

Just in summary, the triage BNP test is a fluorescence immunoassay for the quantitative determination of BNP in whole blood and EDTA plasma specimens. The BNP test device is a single-use plastic cartridge that has a murin polyclonal antibody conjugated to a fluorescence latex particle in the reaction chamber.

A monoclonal antibody that is specific to another epitope on the BNP molecule is immobilized in the detection zone that is read and analyzed by the Triage meter after the reaction is complete.

The sponsor measured BNP levels in three different populations; normals, hypertensives (without CHF), and CHF. These studies were conducted at four clinical sites. Most of the normal patients were obtained by Biosite from apparently healthy individuals in an industrial part setting in San Diego.

One hundred and twenty of the normals were collected at the four evaluation sites. The hypertensive samples were collected both at Biosite and at the four evaluation sites. 93 of the hypertensives were collected at Biosite. Samples from CHF patients were collected only at the four evaluation sites.

Patients were not randomly selected but selected sequentially as they arrived at the clinic. Patients in each of the New York Heart Association classes were studied. The following table shows the number of patients studied based on the data supplied to us by Biosite.

This slide shows the number of patients studied in each group. In the four CHF categories, more men than women were studied, as you can see. The slide shows the numbers for normal men and women, hypertensive men and women, and the four CHF classes for men and women.

This slide gives a summary of the population enrolled in the study. The number of men and women are equal for the normals and hypertensives. As you can see, for the CHF for all classes, only 50 women were studied. Age was not provided on 44 of the normals, eight of the hypertensives, and two of the

CHF patients.

As has already been discussed, BNP levels increase with age and this is going to be further discussed by our statistician Dr. Kondratovich in her presentation.

Next I want to talk about precision for the assay. Biosite conducted its own extensive precision studies on site. The in-house precision was determined using three lots of BNP tests. Each of three controls, low, medium, and high control, was tested 10 times each on 10 consecutive days. Each lot of material was tested using six triage meters.

This slide gives a summary of precision as proposed to appear in the product labeling. The low control used here is 29 picograms/mL, the medium control is 584 and the highest 1,080. Kind of commit the C.V.s to memory for later.

A clinical study was performed to determine if the investigators could obtain the same precision as that obtained by Biosite. Ten replicates of a low and high control were run on three different days at each site. All four produced comparable data to that obtained by Biosite, although no controls were run with levels at the high end of the reportable range as was done at Biosite.

This slide represents the averages for the three days for each site. I put the C.V.s in green so they would kind of stand out for you. You will notice that the low control used here is about the same as was used at Bussed but the high control used here is much lower level than the one in the previous slide that was done at Bussed.

Next I will be addressing interferences with the BNP assay. The sponsor tested the endogenous substances hemoglobin, bilirubin, cholesterol, triglyceride, and the effect of low and high hematocrit. These studies all showed no significant interference with the method.

The sponsor also tested the effects of various of various drugs on the recovery of BNP from blood specimens. These drugs included but were not limited to drugs that are prescribed to patients being treated for congestive heart failure. A variety of over-the-counter medications were also tested. These studies were performed according to NCCLS guideline EP-7.

Two drugs showed interference of 10 percent or greater when added to a 33 pg/mL BNP control. Lovastatin produced a recovery of 91 and 90 percent at 8.00 and 40.00 micrograms/mL, and

Integrilin produced a recovery of 86 and 61 percent at a level of 15.00 and 75.00 micrograms/mL.

That concludes my summary. Next Dr. Kondratovich will give her statistical analysis. Thank you.

DR. KONDRATOVICH: Good morning. I'm Marina Kondratovich, mathematical statistician from the Division of Biostatistics and a member of the team reviewing this test.

I would like to speak about age-matched ROC analysis for healthy normals versus all CHF classes, healthy versus patients with CHF from Classes I and II. Also, I will consider the age-matched ROC analysis for hypertensive versus all CHF classes, and hypertensive versus Class I and Class 2.

The performance of diagnostic test can be described in the terms of ability to correctly discriminate subjects from two groups; non-diseased group and diseased group. We have values of the diagnostic test for the diseased group and for the non-diseased group. Bigger values of the test are associated with disease.

For the given cutoff, a fraction of true negative (green area)is a specificity and a fraction of true positive (orange area) is a sensitivity. A

relationship between sensitivity and specificity of a test over all possible cutoff values is the ROC curve.

The area under curve (AUC) is the good measure of the performance of test because AUC is the sensitivity averaged over all possible values of specificity.

It is well known fact from the literature that the BNP test becomes higher with increasing of age. This is relationship between NBP values and the age of the company data. This is a mean of BNP test and this is a median.

In this situation it is important that the diseased group and non-diseased were age-matched. In other words, one to one matching, for example, means that in both groups there is the same number of subjects for each age stratum.

Otherwise, if we have that the group of non-diseased subjects is younger than the group of diseased ones, then we overstate sensitivity, specificity, and AUC because some difference in the distributions of BNP values is due to the difference in the age. But we would like to measure difference in the distributions of BNP test due to the disease status.

Now we compare two groups; healthy and

patients with CHF, all four classes. The company has 418 subjects in the healthy group, 44 subjects with missing age. Therefore, for age-matched analysis, we have 374 subjects. In the CHF group company has 412 subject, two subjects with missing age. Therefore, for age-matched analysis we have 410 subjects.

We consider such seven age strata; under 25 years old, from 26 years old to 35 and so on. Last stratum is people over 76 years old. This is a distribution of age in healthy group and this is a distribution of age in CHF group.

You can see that the normal group is much younger than the CHF group. Difference in mean is 31 years and difference in median is 34 years. Therefore, these two groups are not age-matched. In this situation sensitivity, specificity and area under ROC curve is overestimated.

We performed aged-matched ROC analysis in such a way; one to one age-matching means that for each particular age stratum we have the same number of subjects in non-diseased and diseased groups.

Therefore, we take all five healthy people in this age stratum and we take some five subjects from this age stratum from CHF group. We take all six healthy subjects from this group and some sick

subjects from this group, all 13 subjects from this group and some 13 subjects from this group and so on. We took all nine CHF subjects from this stratum and took nine from this stratum and so on. That way we receive 84 subjects in each set and these sets are age-matched.

This is the distribution of age in these age-matched groups. This is like we call the effective sample size. In each of our groups we have only 84 subjects, but these two groups now are age-matched.

For this particular age-matched sets, we calculate sensitivity, specificity for the given cutoff, and the area under the ROC curve.

Then we consider the other selection of five subjects from this age stratum, other selection of six subjects from this age stratum, other selection of nine subjects from this other selection, six subjects from this other selection, and three subjects from this age stratum. Again, we receive age-matched groups. In each group we have the same number, 84 subjects. Again, we calculate the sensitivity, specificity for the given cutoff and the area under ROC curve.

Naturally after we consider all possible

variance of age-matching of this data set, we can calculate sensitivity, specificity, and area under the ROC curve as an average of all our calculations.

It is the same as we use simultaneously all observations but with some weights. Observations of healthy group from this stratus, for example, weight one and observations from this age stratum has weight 5/135. For example, observations from this age stratum has weight one and these observations have weight 3/40.

So, this is the distribution of all age-matched groups. Sixty-five percent of subjects in this group from the age of 46 to 65 years. Please pay attention that for this age-matched analysis we used all healthy subjects and all CHF subjects. Therefore, all our estimations of sensitivity, specificity and AUC are more precise than if we use only some one set of 84 subjects from healthy and some one set of 84 subjects from CHF group.

These are the results of age-matched ROC analysis. The red curve is an ROC curve for non-age-matched analysis and the green curve is an ROC curve for age-matched analysis. For non-age-matched ROC curve the sponsor receives this area under curve. This is a confidence interval. Age-matched ROC

analysis gives us set number 0.92. Naturally that for the age-matched analysis the area under curve is smaller but much bigger than 0.5.

In the ROC analysis three values; cutoff, specificity, and sensitivity are connected. If we control one of these values, then we can calculate two others. We consider that specificity is 0.95.

Then in the non-age-matched analysis we obtain cutoff 45 and sensitivity is 0.91. In the age-matched ROC analysis we receive for the specificity cutoff 55 and sensitivity is 0.83. This is the confidence interval.

Now we consider the ROC analysis for healthy normals versus CHF Class I and CHF Class II. We have 209 subjects from first and second classes of CHF. We have the similar picture of the age distributions in groups. The subjects from this group are much older. Difference in mean is 30 years and difference in median is 34 years.

The groups are not age-matched and, therefore, all characteristics of ROC analysis are overstated. Therefore, we performed age-matched ROC analysis in similar way. In each of the subsets we have only 60 subjects but we use all possible combinations. This is the distribution of age-matched

ROC analysis. You can see that about 74 percent are people from 46 to 76 years old.

The results of age-matched ROC analysis are next. The area under ROC curve in age-matched ROC analysis is 0.88 and cutoff 55 gives us specificity and sensitivity. This is the confidence interval. Sensitivity in this situation when we compare only Class I and Class II is 0.77.

In the previous slides you can see that when we compare healthy normals versus all four classes, then we have sensitivity 0.83. Therefore, in this situation we lose about 6 percent in the sensitivity.

Now, let me compare the hypertensive group with diseased group. I decided not to mix the normals and hypertensive group. Because the hypertensive patients usually have bigger values of BNP test, then specificity of the test depends on the proportion between normals and hypertensive in the group normals + hypertensives.

Therefore, in the company data approximately for each four normal patients we have one hypertensive. Specificity for different proportion like, for example, one normal to one hypertensive can be different. Therefore, I decided

if I compare only hypertensive versus all CHF classes, we can receive better understanding of how well this test can work.

Second, you can see that usually people with problem but without CHF has bigger values and statistically some kind of close to hypertensive. Therefore, I decided to compare only hypertensive versus CHF class not combining normal hypertensive versus CHF class.

The company has 167 subjects in hypertensive group. Eight subjects were missing age, therefore, in our hypertensive group we have 159 subjects.

This is a distribution of age in hypertensive and CHF groups. The hypertensive group is younger than the CHF group. Difference in mean is 21 years and difference in median is 24 years. In this situation we again see that this group are not age-matched and we need to make age-matched analysis.

This is the distribution of age in all age-matched groups. 73 percent are the subjects from 46 to 75 years old.

This is the ROC curve for non-age-matched analysis. This is the ROC curve for age-matched ROC analysis. In age-matched ROC analysis area under

curve is 0.87 and this is the confidence interval, much bigger than 0.5.

The results of age-matched analysis are following: In comparison normals versus CHF the cutoff 55 gives specificity 0.95. Now, when we compare hypertensive versus all CHF class, we receive smaller specificity. The cutoff 145 give us specificity 0.9 and sensitivity 0.66. This is the confidence interval.

Therefore, the specificity is very important and then we can choose this cutoff with big specificity. The cutoff 100 gives us more balanced picture, specificity 0.85 and sensitivity 0.75. This is the confidence intervals.

Now let me consider the most difficult situations for the BNP test. Now I compare hypertensive versus CHF for Class I and Class II. Hypertensive is younger than CHF Class I and Class II. Difference in mean is 20 years and difference in median is 23 years.

We performed age-matched ROC analysis. Effective sample size is 78 subjects in each group. 76 percent are patients of age from 46 to 76 years.

The area under curve in age-matched ROC analysis is 0.80. This is the confidence interval.

Of course, sensitivity became smaller if we compare whole group CHF but still relatively high.

There is a difference in the values of BNP tests for the healthy men and healthy women, hypertensive men and hypertensive women. This is the female normal, hypertensive. This is male normal, hypertensive. You see that an average woman have bigger values of BNP test.

This is box-and-whiskers plot of values BNP test for females and males. This is the males normal, females normal, males hypertensive, females hypertensive. Yellow box has 50 percent of observations.

This is box-and-whiskers plot for CHF subject. Naturally this observation from these four classes are very overlapping.

It looks like there is some difference in the values of BNP test of CHF male and CHF female. A woman on average have bigger values of BNP test. You can see that, for example, for Class I male 118, female 138. Male for Class II 310, female for Class II 555. Male for Class III 701 and female for Class III is 811. Male for this class we have 1,526 and female we have more than 2,000.

We cannot consider this picture as very

reliable because there are very small number of observation for female compared with picture when we see difference in normal group between males and females and in hypertensive group between males and females.

This picture is very reliable because we have relatively big number of observation for women and men. In this situation this observation for female is relatively small, only five. When I performed age-matched ROC analysis for normal females versus CHF females, I receive that area under ROC curve is 0.88.

Of course, confidence interval is relatively wide because of small number of observations. Lower limits of this confidence interval for area under ROC curve is 0.73.

This is the box-and-whiskers plot for females and males with CHF. You see that this is the male/female for Class I and male/female for Class II, male/female for Class III, male/female for Class IV. This line is the median.

You see that women have tendency to have bigger values of BNP test. Therefore, gender can contribute only to potential misclassification of CHF subject because naturally there is considerable

overlap between CHF classes.

Also precision which is 12-16 percent and drug interference can contribute to potential misclassification of CHF patients but do not affect significantly the ability of the test separate CHF patients from others. I mean that normals from CHF and hypertensive from CHF.

Thank you for your attention.

DR. KROLL: Thank you very much.

MS. CHESLER: I would just like to summarize for the group the questions that we have for the panel. (1) Using 55 pg/mL as the final cutoff resulting in the following performance parameters: Age-matched healthy controls versus all patients with CHF; sensitivity 83 percent, specificity 95 percent. Age-matched healthy controls versus patients with CHF (Class I and II); sensitivity 77 percent, specificity 94 percent. Is this the appropriate cutoff or should it be raised or lowered?

(2) The study design was a model studying a pre-selected population (healthy controls, hypertensives, and patients with defined CHF). Although results closely approximate sensitivity and specificity reported in the literature, the test was not studied in actual emergency room use. Should this

be indicated in the labeling?

(3) There are several ways to portray the data and to calculate sensitivity and specificity. These could include comparisons of healthy controls to all patients with CHF, comparisons of healthy controls + hypertensives to all patients with CHF, or similar comparisons to early states (hard to diagnose) CHF. What should be included in the labeling to ensure that users understand the potential variable performance of the assay?

(4) FDA has evaluated the cutoff using age-matched data and ROC curves. Is this the appropriate analysis? Do you have other suggestions on how data should be analyzed and presented?

(5) There is considerable overlap between the NYHA CHF classes, and FDA is concerned that gender differences, assay precision, and drug recovery can contribute to additional overlap or misclassification. Should the BNP results stratified by the NYHA classification remain in the labeling as is, be modified in some ways you could suggest, or be deleted? Thank you.

DR. KROLL: Thank you very much, both of you. I think in the interest of time we are going to break for lunch now and then we will come back in an

hour at 1:45. At that time the panel can then address questions to both the FDA speakers and to the sponsors and we'll try to address these issues you've given us here. Thank you.

(Whereupon, off the record for lunch at 12:46 p.m. to reconvene at 1:45 p.m.).

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

(1:46 p.m.)

DR. KROLL: Good afternoon. I hope everybody has recovered from lunch. We would like to get started again. Now we're going to start with the committee discussion. The first thing we're going to do is for each of the panel members whether they have any questions for the two FDA presenters.

I think we need to have that table clear. You can leave the stuff, just if you would get up because we want both FDA presenters to be able to answer any questions.

The first person I saw was Dr. Rifai.

DR. RIFAI: Yes. I have a couple of questions regarding the statistical analysis. It's the logical thing to do since the concentration various with age and with gender is to correct as you have done. The only problem that could create is the number of observations become relatively small.

I wonder if you have done retrospectively power analysis to see that with this number of observations you have you can reach good statistical power?

DR. KONDRATOVICH: Yes. For example, consider the problem of age. I divided only for three

groups like relatively young. Therefore, it's very natural that we can have different cutoffs depending on the age group and the different cutoff for different gender. If we can use bootstrap, then we can estimate cutoff and sensitivity specificity relatively good.

I made age-matched ROC analysis of only separately for normal women versus CHF women. I obtained that area under the curve in this situation is 0.88 and confidence interval lower limit 0.73 and upper limit is 1. Then the test work in this situation.

Different problem how we need to evaluate cutoff but there are techniques like bootstrap. These techniques can give relatively good results even in the small numbers. If company makes additional analysis and make more observation because, for example, it's very difficult to make any age-matched analysis.

For example, even for normal women we have in the age 56 to 65 only four observations. In the age 66 to 75 we have only two women. Over 76 we have zero observation. Therefore, age-matched analysis make little bit difficult to make. Not enough observation for women.

But I understand that this is connected with congestive heart failure that women usually have less. Smaller number of women has this disease so there is a small number of observation in CHF. This is for normal women. I think company can have more observation for normal and for hypertensive.

DR. KROLL: I think also Dr. Packer has a question.

DR. PACKER: Just for the record, if you look at the entire population, heart failure is more common in women than in men. It comprises about 50. If you look at all patients with heart failure, 53 percent are women, 47 percent are men. The reason you don't get that impression in clinical trials is that clinical trials use a cutoff of a low ejection fraction.

Whereas women primarily have diastolic dysfunction, men primarily have systolic dysfunction. For clinical trials we have 80 percent men, but if you look at the entire population it's really a disease that occurs slightly more commonly in women. The lack of data here in women is actually quite important.

DR. KONDRATOVICH: Yes, you are absolutely right. In reality this study is like only male study because there are in CHF group most of observation is

male.

DR. PACKER: I want to ask one question. Before coming to today's meeting I asked one of the statisticians at the university about what could be done to adjust an analysis if there's an imbalance in age which is what you have tried to do.

He mentioned a number of things but one of the things he mentioned was the approach that you have taken which is to try to select out from the group that was studied and try to match equal numbers of patients. I would honestly say that he had significant concerns about that approach.

DR. KONDRATOVICH: Because I always show you the distribution of age in my age-matched group. This age distribution must reflect the age distribution in the target population.

DR. PACKER: I think you probably have hit the nail on the head. The problem with what you have done is that when you have an age-matched population and there are very few elderly people, then there will be very few elderly people that are matched. Consequently, when you construct your ROC curves, you will be constructing your ROC curves on a patient population with a distribution of ages which include very few elderly people.

DR. KONDRATOVICH: Yes, you have said it right.

DR. PACKER: This is a good representation, I think, of the problem. The ROC curves that are age adjusted, the age adjusted ROC curves are weighted according to the ages that you see here. The problem is that the amount of weight in the patient population that is relevant here over the age of 65 is an 18 percent weight.

DR. KONDRATOVICH: Yes. You have said it right.

DR. PACKER: But the ROC values are constructed on the whole population so if you had terrific sensitivity and specificity for young people and terrible specificity and sensitivity in old people and you tried to age adjust the ROC curves using this distribution, you would only discover a small drop off in the ROC value because a number of elderly people that contribute to this analysis is so small.

DR. KONDRATOVICH: Yes, you have said it right. Therefore, it's very different from this distribution. In this situation you need to make a separate analysis for very elderly people, middle aged and very young.

DR. PACKER: Right.

DR. KONDRATOVICH: This analysis will give you some kind of average.

DR. PACKER: How comfortable are you that the age adjusted ROC values mean anything?

DR. KONDRATOVICH: It means that if we consider that in one set we have sad distribution of age, you need to verify this distribution of age reflects some distribution of age in target population. If you say that I don't agree with the distribution, then our data set cannot give us this information.

DR. PACKER: That's exactly right.

DR. KONDRATOVICH: For example, you can say that I am very interested in the ROC for very old people. Then I need to have different distribution of age.

DR. PACKER: I don't think we're interested in the ROC of old people. I think we are interested in the ROC of people with heart failure. It just so happens that people with heart failure are old. This would be a valid analysis if this were the distribution of age in people with heart failure.

DR. KONDRATOVICH: Yes.

DR. PACKER: But this is not the distribution of age in people with heart failure.

What we want to see is a ROC age adjusted analysis where the age distribution resembled the disease for which the test is being proposed.

DR. KONDRATOVICH: You're absolutely right.

DR. PACKER: From what you've told us, the data are insufficient to construct such a curve.

DR. KONDRATOVICH: Yes. Then in this situation statistics allow us to receive all information that we can obtain from this data set. In this data set I cannot make analysis for old people or maybe the same for very young people because of the same problem.

DR. PACKER: I understand that. Again, the ROC values based on this distribution are not the ROC values for the use of this test and the target population, even age adjusted because this is not the age distribution for the disease. Therefore, I always pay attention to how many percents at what age.

DR. PACKER: Right. Very small. Here is the reason for worrying. If one uses, for example, a cutoff that your analysis proposes which is 55 picograms/mL, and one looks at the graph on page 252 in Volume I, and one looks at patients over the age of 60, not 65 but over the age of 60, it's very

interesting.

This is the total number of patients in the database submitted in the application in patients over the age of 60. Total number of dots here is 13. 14 if you count the borderline one. If you use 55 as the cutoff, then seven -- no, actually it's eight if you include the dot on the borderline -- eight of the 14 are false positives. Eight of the 14 are false positives.

That means that a patient population at risk, which is the patients who are over the age of 60, an elevated value above 55 is more likely to be consistent with normal than it is with heart failure.

DR. KONDRATOVICH: Yes, you are absolutely right. In this situation when some test depends on age, it's very good to have different cutoff for different group age. This data --

DR. PACKER: It sounds as if we can only develop a cutoff for this test for people who don't have the disease in an age group that is not at risk.

DR. KONDRATOVICH: This is 55 cutoff. This is cutoff between normals and CHF.

DR. PACKER: I know but that's the primary purpose that the test is being developed.

DR. KONDRATOVICH: I understand that maybe

not right now hypertensive reflect more.

DR. PACKER: It would be worse if one did this in a hypertensive group.

DR. KONDRATOVICH: Yes.

DR. PACKER: It would be worse. If one did it in the hypertension group and the number of false positives using a cutoff of 55 would be 16 out of 23.

DR. KONDRATOVICH: Yes. Therefore, for hypertensive we need to make bigger cutoff.

DR. PACKER: Maybe we need more data.

DR. KONDRATOVICH: You mean hypertensive with a depend from age, then when I make a cutoff for a particular hypertensive versus CHF, I consider that I can make only age match according to this age distribution with what I receive from this data set.

DR. PACKER: There are only 13 normals, 23 patients with hypertension in the control group with confidence intervals from here to Capitol Hill.

DR. KONDRATOVICH: For example, here this is hypertensive versus CHF and if I consider the distribution in my age match --

DR. PACKER: But this is not the distribution for the disease. Your ROC values are

terribly weighted by the fact that you are -- this is a very good test for young people so that by including a lot of people under the age of 60, you are shifting your ROC values upwards. You are overestimating the true ROC in the patient population at risk.

DR. BRINKER: I think the problem here is that --

DR. KROLL: Excuse me. Can you state your name for transcription purposes.

DR. BRINKER: Jeff Brinker. I think the problem here is that the statistician is trying to do the best she can do with the numbers given and not looking at the alternative which is to get better and more numbers. I think she would probably agree that if you had all the information available, you could do the right status.

DR. KONDRATOVICH: Yes, right status. Always pay attention and do you like this distribution and do you like these numbers. Do these numbers reflect some kind of picture what you would like to see in age match. If you say that no, I don't like. I would like to see more age. Therefore, statistics cannot help, all information from this data set. This data eliminate the fact of age. Now I see only the effect of disease status but, of course, for this age

distribution.

DR. KROLL: All right. Thank you. Are there any other questions for the FDA presenters?

DR. KONDRATOVICH: Excuse me. You see that I make one group like over 76 years old because in company data there are a lot of people like 80 or 90 years old but there is no observation for normal. Therefore, in this situation I can make only one group but, again, if I merge this group, there are some biases in this situation because right now I consider that they all are the same age, but in reality no.

DR. KROLL: I believe Dr. Clement has a question.

DR. CLEMENT: I have just one brief question. This is regarding looking at the standard deviation C.V.s that were done on the precision data. You had mentioned that you would come back to that at some point in your presentation. I don't have a good feel except for looking through the articles that were submitted to us what a good C.V. is for this type of test. I mean, we think of C.V. done in the laboratory situation as being 2 to 3 percent in in vitro method.

MS. CHESLER: That's very dependent on the type of assay, of course. I did go back and look at the data on the Biosite's previous device which is the

cardiac panel and these C.V.s seem to be pretty close to what they had for that panel test.

For this device that looks like it fits in with what you're going to get with this device. Of course, it's going to be a much higher C.V. than if you had something right in the laboratory and it was sodium or something like that. It isn't fitting with the device that is already cleared and in use.

MR. REYNOLDS: I have a question along these lines.

DR. KROLL: State your name.

MR. REYNOLDS: Stan Reynolds, Consumer Rep. Along these same lines there was a big difference in the mean value of the high control that was used by the --

MS. CHESLER: Right. That's true. So you can't really compare the C.V.s directly because the control used was so different. Of course, I think the values obtained by the four evaluation sites actually had lower C.V. but that's probably because of the controls they used were different. I think for the total precision of the assay, I think what was provided by Biosite is probably something better to judge it by.

DR. KROLL: I think Dr. Brinker has a

question.

DR. KONDRATOVICH: If we have this age distribution by the same weight we can -- I mean that by special selection of weight for this particular group we can receive estimation of the ROC curve and sensitivity and specificity for distribution what you like.

But, of course, the confidence interval will be much, much wider. Therefore, I decided not to do this and use in optimal way this data. If I put more weight for this age particular, then my confidence interval will be bigger. It means that I need more observation here.

DR. KROLL: Thank you. Dr. Brinker.

DR. BRINKER: Just a quick question to the FDA staff. Are there other commercially available FDA approved assays for any of these peptides, BNP or AMP?

MS. CHESLER: There are not.

DR. KROLL: Dr. Everett has a question.

DR. EVERETT: This is to the FDA staff. My question is the way the test of proposed to be used at this time, does it seem to have -- age doesn't seem to make a difference in terms of how it's being proposed and that is the reason you went back and did the age adjusted data. Is that correct?

DR. KONDRATOVICH: Yes. Yes, absolutely correct.

DR. EVERETT: And in this sense since hypertension and congestive heart failure is a real issue for African-Americans and they do use some African-Americans in their data, did you look at that to see if it was actually safe to be used in that group?

DR. KONDRATOVICH: You mean separately for African-American?

DR. EVERETT: Right.

DR. KONDRATOVICH: Our company has information about race but because of small number of observations, therefore, I did not make separately this analysis for particular race.

DR. EVERETT: So no statistical support for using it or not using it in African-Americans. Is that correct?

DR. KONDRATOVICH: I did not make this analysis separately. Therefore, in all this statistical analysis race I did not consider.

DR. EVERETT: Okay. So as a safety issue then, which groups statistically would you say they evaluated well enough that statistics would support using this particular device in that group of people?

DR. KONDRATOVICH: Statistically concerning the gender because the most subjects in CHF are males. Therefore, this is like analysis more for males, not for females. We can only estimate the area and the ROC curve for females and see that this is area more than 0.5 and very difficult to make more precise conclusion for females.

Concerning the race, I think that this test reflect like some kind of average structure of American society because in their population Americans have some kind of representative sample data set. Of course, not enough observation to make separately study. Therefore, this is like some kind of average male.

DR. KROLL: Excuse me. This is a time for the panel to be raising comments and other questions and concerns. Are there any other specific questions about things that the FDA presented to the panel before? Because we're going to have some time now if there aren't any of those particular questions to go ahead and raise other comments and concerns.

DR. HENDERSON: I'd like to follow-up on --

DR. KROLL: Cassandra Henderson.

DR. HENDERSON: I'd like to follow-up on

Dr. Everett's question. Is the group large enough if you look at African-American males with hypertension to look at them as a separate risk group that might benefit from this device?

DR. KONDRATOVICH: I did not consider this statistical analysis.

DR. GUTMAN: Can we defer that question to the sponsor actually?

DR. KONDRATOVICH: Yes. You're absolutely right.

DR. KROLL: We will come back to the sponsor to answer that question because I want to spend some time now for each panel member to make comments or raise other concerns that they have and initially address a question to the sponsor. Then what I would like to do is have all the panel members be able to speak first and then we'll come back and revisit those questions to the sponsor.

The sponsor will have time to actually hear the question first, think about a response, and then we can try to address all those questions at the same time rather than coming back to those questions from each person.

Since I'm Chairman, I'm going to just mention a few concerns I have right now and then we'll

go to Dr. Rifai. My concerns, at least ones that haven't been brought up so far, have to do with some of the analytical techniques that are involved. One of the things I was impressed with in the packet that was given us is that there were a lot of articles to previous studies in literature.

But there is a problem and that problem is that I didn't see anywhere in the packet where there was a comparison between this proposed method and those other methods that are out there in literature especially dealing with peptides and proteins. It's difficult for me to have much confidence unless there is some type of comparison.

The other aspect of that is we're talking about peptide. I would have liked to have heard some information about standards that are used. Is there a standard preparation of the peptide available? How is the company going ahead and establishing their standardization?

How are they linking that back to the actual products that they produce so that we know that each component is actually linked back to a standard so that when you measure it today or you measure it two years from now and you get a number of 50 that is the same number. That information is not available.

At least, I didn't see it.

That raises another issue which is I did not see presented frequency for calibration verification, the types of materials that should be used, how that should be done and how often. That's an important point especially for any method being introduced is that it should be well characterized in terms of doing calibration verification.

I did not see discussion of how the calibrators were made or what are in them and how they should be used. I didn't see information for the stability of control material nor for calibrators.

Additionally, I did not see any studies firming the liniarity of the method or how to handle samples if you are above a certain amount. That's another critical area where we have missing pieces of information.

Let's see if I have anything else that's on my list. Oh, one other thing in terms of interferences. This is a method that is based on a fluorometric approach. Historically if you look back at methods based on fluorometric approaches, and I can think of one company that has had a method available for a long time, they have had problems with samples with people who are in renal failure.

It's not necessarily because they are retaining something related to the analyte but they are retaining materials that fluoresce. Serum has natural fluorescence and sometimes these patients have very high fluorescence. They interfere in two ways. One, they actually interfere directly with the assay or, second, they provide a tremendous amount of background and it interferes with the method that way.

I did not see any assessment of those types of issues or where they went ahead and tried to find samples from people who have this and go ahead and try to characterize the amount of fluorescence that was inherent and when there is high background fluorescence and see if that had any affect in their assay.

Those are the rest of my comments. The sponsors can hear that and they can try to address them later. Let me go to Dr. Rifai.

DR. RIFAI: I had several concerns and I'm just going to mention the ones that I have on my list that were not mentioned by Dr. Kroll. In regard to the precision study, one thing that I thought was quite interesting, usually the imprecision get better as the concentration goes up. In this particular device it is the other way around. I don't know what

explanation for that. The sponsor might have an explanation. This is just an observation.

I felt the description of how the studies were actually done was not presented clearly. For example, it is not stated if it's the same person who does the precision study from day to day or different people to reflect real life situations. This is one thing.

The other thing, the sensitivity was presented was the analytical sensitivity which is taking basically the buffer and plus/minus to standard deviation and would be important to know the functional sensitivity with actually the device is capable of measuring.

Again, this device was demonstrated to be used in whole blood or EDTA plasma. The question is is there a reason why heparin and serum were not -- is that because the sponsor has not looked at that or because they examined the heparin and they found it, for example, interfering with the test? These are important things to bring out.

Dr, Kroll mentioned briefly on the calibrator. I went a couple of times trying to figure out how actually the device was calibrated and there is just a very simple description in passing that the

pure compound was used. Since here you have a built-in calibrator, I think it would be very important for us to figure out how you came up with this internal calibrator and this information was not provided in this application.

Some of the logistics of the clinical trial I felt were not very well described. I mean, a lot of information was provided but, for example, the samples were collected and some of the samples were split and sent to Biosite. Others were measured on site in one of these clinical sites. It was not indicated, for example, for your calculation which values did you use.

Did you use the ones that were generated by the site or generated by Biosite? How blinded were the investigators since they know which one has congestive heart failure and which one has the controls? Who did the analysis where the data from the BNP data and the clinical data were married together? Were they done by the individual investigator? Just some description about how the actual study was done will be very beneficial for us.

Again, it wasn't clear if whole blood or plasma was used. I think it was plasma but, again, it would be nice to confirm that. From the presented

data for the LV studies it seems like really all the data presented, or the great majority of them, involve men so whatever claim is going to be made, if it were to be made, that should only be directed to men and not to women. It's a far stretch to extrapolate that to women without data.

My main concern besides the issue of controls that has been brought up before was the location of where the actual analysis of BNP was performed because this is a test that supposedly will be very helpful when it's done in an emergency department. In these particular studies this test was done in a laboratory environment by laboratory personnel.

If we learn from history about the performance of New York patient testing, we will know that the instrument no matter how simple they are, the quality of testing is different when you take it from a controlled environment of the laboratory. This is really something that the sponsor must address.

It was mentioned in passing that some work has been done but certainly none of that work is included in the materials that were reviewed. That's it.

DR. KROLL: Thank you. Please state your

name, sir.

DR. ROSENBLOOM: Arlan Rosenbloom. I wanted to specifically address the diabetes subpopulation which is a substantial one comprising of a fourth to a third of all patients in this group. My understanding -- remember I'm a pediatric diabetologist -- my understanding of myocardial insufficiency in diabetes is that it's a somewhat different disease both in terms of symptoms that patients have as well as the findings.

I wonder if the New York criteria are specifically applicable to this subpopulation if the age criteria are specifically applicable since people with diabetes have a form of accelerated aging. Also the metabolism of the BNP may be somewhat different for some of the reasons that have been stated; renal involvement in diabetes and the presence of fluorescence substances, particularly collagen products and other glycated proteins in the circulation.

I think that we need an analysis of the diabetes population to see if as with other populations that have been mentioned, African-Americans and women. Thank you.

DR. KROLL: Dr. Henderson.

DR. HENDERSON: I don't really have any other questions other than what I asked before, just to ask if there was a way or if you had the numbers to analyze African-American males who had hypertension as a separate group to look at the predicted values of this device.

DR. KROLL: This is Dr. Kroll. I want to enlarge on that question a bit. I think that the question we talk about ethnic and racial groups, we should think beyond Caucasian and African-American but also think in terms of people of Asiatic ancestry, Native Americans, and other major pertinent groups which may actually release this peptide differently than major groups or the group that's been studied where there may be significant differences. It doesn't have to be a huge study but has to be done in sufficient numbers so that we can draw a conclusion.

DR. HENDERSON: I agree. My only concern was in looking at the numbers that they have, since we have the limited data that we have of elderly people and also other groups, that I think it's facing us immediately with the data that they have.

Perhaps the largest group that they might be able to look at whether there is some validity in

the predicted value and is that population particularly at risk for hypertension mentioned. I absolutely agree that other ethnic and subpopulations are very important. In the future we need to collect more data.

DR. KROLL: Dr. Brinker.

DR. BRINKER: Thank you. I think I can reflect what most of the people here feel, that this is a far from optimal database upon which to make certain decisions. However, I don't think that would preclude approval of the device.

What I would look for in wanting to approve a device, even if we don't have all the information because some of that can be garnered later, is, No. 1, does the assay reliably measure the peptide? No. 2, is the level of the peptide in blood reliably associated with the pathophysiologic parameter of interest?

In this case, I believe that is LVEDP basically, filling pressures. This doesn't mean it has to be directly related but relatively related to it. Enough to give some information about the patient. This information could be gathered by alternative methods but if it is reliably monitored by this method, that would be, in my mind, a reason for

approval.

Is the level of the peptide affected or the assay affected by other metabolic processes or pharmacologic manipulations? What is the likelihood of misinformation from this assay causing patient harm? What more information does the sponsor need to do to optimize the data set, not necessarily now but to better direct how to use this assay in the future?

I think there are certain minimal pieces of information that are necessary for approval. I don't think you need to show that this reflects EDP exactly. I don't think you need to show that there isn't some overlap between severe heart failure, minor heart failure, gender, and age.

What I think you need to show is that there is some relevant association between your assay and the parameter and your assay and the peptide and the peptide and the parameter that it is supposed to reflect. To be honest with you, the market will determine how good this assay is in the future.

DR. MANNO: Barbara Manno. I have some questions on the device itself because we've not touched on that here. I'm trying to rationalize some of the differences in numbers based on the performance of the device itself.

What I'm referring to, I have no idea how the device is calibrated in actual use, how many calibrators are used, and how often it has to be calibrated. Those things would very definitely impinge upon the within-day meter performance as well as the device performance itself.

I'm also curious what is an industry, let's say, accepted C.V. for performance within-meter performance, in between meter performance. I know what I use in my own lab but I don't know what might be in this type situation. That would be nice to see something on that.

And we haven't heard anything about what type people are doing the within-meter and between-meter days and within-day and between-day studies done by the company if this data is coming from company personnel versus the clinical sites and how that might be looked at whether we're talking about nurses, just ordinary chemists, med techs, whatever.

I know in most laboratory environments based on the training of the individual, the classical training of the individual will give you -- a med tech will give you generally narrower C.V.s than will someone from a general chemistry background. This is an example only. I'm not criticizing the chemists by

any means or lauding the med techs.

I also am a little curious when you're talking about in the manuscript, especially on page 246 of Volume I. We've pretty well established, I think, here that we don't have enough women looked at in this study, or we don't have very many.

Yet, you're saying that your data agrees with the literature and you give a citation here that the difference that you see between men and women is not due to menstrual cycle, age, or any factor that can be identified.

That brings me to interfering substances. I would be interested if you add women to know what happens to those who are or are not on estrogen supplements, whether that makes a difference or not in the actual values that you would establish as normals.

DR. HENDERSON: If you are adding women who are on estrogen, then --

DR. KROLL: Remember to identify yourself.

DR. HENDERSON: Cassandra Henderson. That would be an older population.

DR. MANNO: That's why I'm bringing it up because there are a number of patients in our hospital that won't use the estrogen supplements so you might have a difference there, even though we're not sure

yet what that is doing in terms of cardiovascular in the older population.

That's my understanding and I agree very strongly with the things brought out by the Chairman in terms of method comparison. While we don't seem to have a method in this country, there are at least two other approved methods in the world supposedly according to the documents supplied. It might be nice to see what a comparison would be against those where they already have been used and we've been able to get data points gathered. I'll stop with that.

DR. KROLL: All right. Thank you. Dr. Gutman.

MR. REYNOLDS: Reynolds.

DR. KROLL: I can't see without my glasses on.

MR. REYNOLDS: Stan Reynolds, Consumer Rep. Being the laboratorian in the group, I pretty much have some of the same concerns that have already been voiced, particularly about the calibration and quality control procedures.

I'm particularly troubled with the precision study because it looks like we're comparing apples and oranges with what was done in-house versus what was done by the four other sites because, you

know, their samples are very different ranges.

I'm not really sure again who did the studies, if it was different people on different days, the same people, if this is something which is going to be used in a ER setting, a point-of-care type setting, and were they always done on the first shift or were they done on different shifts.

I just have some general questions basically about calibration, quality control, and precision issues and I would like some clarification on those.

MS. AMMIRATI: Erika Ammirati. I don't have anything to add.

DR. EVERETT: James Everett. I just would like for the sponsor to address the issue that I have discussed with the statisticians and that is if we're going to be here to determine if something is safe and effective, then we can't say we just do it in one group and now it's applicable to everybody. That just isn't scientific at all.

In reality, there are barriers to trying to do everything and everybody who might be susceptible or exposed to this particular device. My real problem deals with the fact that the people who are most likely to be exposed to this particular

group, African-Americans, Caucasians, not to exclude any other groups.

When I asked the question to the statisticians, regardless of race or gender, which group does the data support that this particular instrument might be safe and effective, they didn't say everybody. They simply said males. They didn't say Caucasians. They didn't say African-American. They didn't say Hispanic, Japanese or anybody. They simply said males.

In a real sense if it's only effective, or looks like it might be effective, in that particular group, then the question is is anything else effective or should we be evaluating it or should we say, okay, we'll figure that out later? That really isn't scientific.

I would like to know, in short, whether the sponsor agrees with the conclusions drawn by the FDA statisticians.

DR. CLEMENT: Steve Clement. After looking at the data, I'm definitely impressed on the age issues, particularly looking at the hypertensive "normals." Again, we don't have the data broken out like that. We're doing our own little mini statistics and actually counting points here.

Basically for folks over 60 years of age that are hypertensive, non-heart failure patients, there is 27 points that we counted. We use a cutoff of 100 picograms/mL as the normal cutoff which is the high end compared to all the proposals there. Ten out of those 27 "normals" have a number greater than 100.

If I think back to the ER situation where it looks very tempting to use this device, of 30 people that come through there, there is almost a 50 percent chance that I would get the wrong decision just based on using 100 as a cutoff. I could almost flip a coin, at least based on the data.

It may be very good but I think as many of the other panel members said, we need more data in that age range of people, particularly in different subgroups so we know what that value means and we know what a normal value is in that group.

DR. PACKER: Milton Packer. I've already raised concerns about the nature of the controls and the selection of patients with heart failure. I think that the concern that I have is that we can really construct ROC curves not just for men but only for young men.

Those are the patients that we can construct a ROC curve for, young men, but young men

don't get heart failure. Old men get heart failure. Old women get heart failure. This is a disease which is predominately a disease of the elderly. That's where this database is most efficient. It's most efficient in the patient population that is most prevalent in this disease.

I appreciate the sponsors saying that this is not intended to replace echo and echos will still need to be done on these patients for all the usual indications that an echo is done. It's hard to imagine, however, how this test will add incrementally to the test that would already be done.

How would it add incrementally to an echo? I'm not certain that it does. I'm not certain there is anything that this test provides in the clinical decision making process or in the diagnosis of heart failure that already wouldn't be provided by tests that would already need to be done in this patient population.

I think there is a danger that the echo won't be done. There are many primary care physicians who avoid sending patients for echocardiograms because they need to generally send them to a cardiologist and they are afraid of losing them after the patient is referred to an echocardiography lab and that they will

rely on this test as a surrogate for an echo. That would be a big mistake.

I think what I really would like for this test to do, and the sponsor has proposed a number of really good ideas that I think need to be explored, but certainly these are not ideas that we have data for that would guide either diagnosis or therapy.

What we really want to be able to do is to take a patient who is 65, 70, 75 years old, male or female, who comes in with shortness of breath and who has all the usual characteristics of old people. They've got a little coronary disease. They have a little hypertension. What you want to know is that shortness of breath due to heart failure or not.

Right now the problem is that I don't have any data to know how I would use that test in the most prevalent example that I would like to use the test. Most importantly I don't know what the cutoff is in that patient population. Steve already said that, gee, you know, you could use 100. It's not very good at 100. You can use 300, 2 data points over 300. I wouldn't have any idea.

I note Jeff Brinker said he really thinks as long as it measures what it's supposed to measure,

that it might be useful, but I don't know what a normal value is. I don't know a value that distinguishes heart failure from patients without heart failure in the patient population at risk.

What I'm really worried about is that we're going to by approving this device create a disease called elevated DNP disease. Just like we have a disease called PSA disease and we're going to drive a lot of people nuts and result in a lot of workups of old people who are pretty just normal old people and who have elevated BNPs because old people have stiff ventricles.

I think that this has more likely to create the impression of disease where there is no disease than to assist in the diagnosis of real disease.

DR. COMP: Philip Comp. I have a safety concern on the part of hospital personnel. When you do a finger stick glucose you stick the finger and measure it. If you use a point-of-care coagulation device, you draw blood in a syringe uncoagulated, put it in the slide and put it in the machine. This is a little different.

Now you've got to draw blood into an EDTA, I assume, vacuutainer, hopefully invert it a few

times, but then somehow get the blood out. In a busy emergency room that worries me. Are you now going to use another syringe to go through the cap of that thing and try to suck some blood out or are you going to pull the cap off and aerosolize that patient's blood in the face of the operator? I don't know.

This is quite prime time in terms of that safety issue. I would like to see that very definitely addressed. I'm not sure right now hospital safety committees would go along with this technology unless it's a little more clearly stated.

DR. KROLL: All right. Thank you. At 3:00 we have to have the open public hearing. What I would like to do now is go around the panel again and people can restate their questions that they would like to hear a very succinct response from the sponsor. I remind people that when they do come up to respond to tell us who you are.

The main points that I brought up before had to deal with calibration, calibration verification, and liniarity. I wonder if the sponsor has some comment to that, whether that has been worked out and where the information is available, and how would it be included in a package insert.

DR. BUECHLER: Ken Buechler, Biosite. The

issue of the calibration. Why don't I start with the standards. Should I try to answer some of the general questions that have common answers or should I answer specific questions?

DR. KROLL: It depends but in a sense like this you might want to respond that you've done the studies, you have the work, and it just wasn't included in here and it may not have been included in the package insert. We don't have to hear every specific point.

DR. BUECHLER: Yes. Okay. So for the standards used, this was purified BNP. This was BNP made by SCIOS. Actually made by Abbott Laboratories, I believe, or SCIOS, one of the two. Abbott Laboratories developed the lyophilized material. We did extensive analysis of that material using maldetoff mass spec and verified that the peptide was greater than, I believe, 95 percent pure.

All of the standards were made that were used in the calibration of the test by weight. Originally the material then was weighed out based on the purity of the material into plasma samples.

Interferences. You had also asked about interferences with other flurometic approaches. The fluorescence that is measured by the instrument is

excited at 670 nanometers and it's emitted at 760 nanometers. This is in the near infrared.

We've done extensive studies indicating that nothing that we know of that's in the plasma absorbs up there in the near infrared part of the spectrum suggesting that there are no optical interferences or fluorometic interferences. Did I answer your questions?

DR. KROLL: Most of them. Why don't we go to Dr. Rifai. Did you have a specific question you wanted them to answer now?

DR. RIFAI: I don't know. Probably the sponsor has already taken notes on what each one of us has asked so why don't we do it that way and we'll save some time.

DR. BUECHLER: Ken Buechler still. In the case of your questions, Dr., you asked about precision studies and why the C.V.s increased as concentration went up. That's atypical of assays and you are correct.

The reason for this is that the dose response curve near the top end of the range the slope of it decreases rather than staying constant. For any relative shift in the signal there's a slightly larger shift in the concentration and that's the reason that

the C.V.s increase with higher concentrations. It's a pure analytical reason.

Who did the precision studies in the lab? Those studies, at least at Biosite, were done by a whole variety of technicians and scientists. There was no special design except that it was generally random in the people who performed the test.

The analytical sensitivity, the MDD that was measured, is an MDD that is measured by the standard laboratory practices that all manufacturers of immunoassays follow, and that is to measure zeros, calculate the standard deviation of that measurement, multiply it by 2, and that response then relative to the dose response curve is the concentration or the analytical sensitivity. We use standards that the industry uses to do this calculation.

The whole blood is EDTA. EDTA is used because this is a peptide that can be protealized in whole blood without EDTA and so the peptide is more stable in EDTA blood for that reason.

The device is calibrated using standards, as I mentioned earlier, that are weighed out. I believe there were more than 10 and I believe closer to 15 standards that were used to generate the calibration curve. Again, the calibrators were all

made by weight.

I think those are all your questions, Dr.

DR. RIFAI: I think I have one general question about some of the logistics of the clinical trial. Can you clarify which of the BNP values were included? Are these the ones who were done at the particular site where the clinical trial was taking place or the ones that were measured at Biosite?

DR. BRUNI: Many of the apparently healthy people were measured at Biosite and collected at Biosite, whereas the patients that were diseased were collected at the clinical site. None of those were measured at Biosite.

DR. RIFAI: Okay. And you used the plasma and not the whole blood? Is that correct?

DR. BRUNI: We used primarily whole blood at the clinical site as opposed to anything that was done at Biosite. We had it identified at Biosite. We had to resolve a discrepancy and we didn't resolve anything.

DR. RIFAI: And can you just comment a little bit about some of the logistics in terms of when the patients were referred or were diagnosed by congestive heart failure at Stage II or Stage III and then the samples were sent to the laboratories. Where

were these two data sets merged? Were they merged at Biosite or merged at a third party?

DR. BRUNI: I think Rob Christenson is probably in the best position to do that since he was one of the study sites.

DR. CHRISTENSON: Hello. Rob Christenson from the University of Maryland. I don't have any financial interest in Biosite.

The logistics of it were that there would be a heart failure clinic at least at one of the four sites. We had a heart failure clinic where we would send a medical technologist to first ask the patients if they were interested in being in the study and then get the informed consent and then to collect the samples there and actually do the test right on site.

Now, the test were performed by a medical technologist. As far as where the data were merged, that was at Biosite so we had case report forms and the medical technologist would record the data on the form without knowing specifically whether that patient was one, two, three, or four but just that that patient had come to the heart failure clinic so it was a high prevalence group.

I guess just to comment on a couple of other questions, one being functional sensitivity. I

think ideally all tests would do functional sensitivity but particularly the ones where it's very important to do functional sensitivity our tests with the low-end values are very important. We can think of examples.

Chiponin might be an example where the low-end values mean something. Certainly TSH is another very important one to define generation, but a test where the low-end is not really where you're focusing where the cutoff is higher, the functional sensitivity becomes a much less important issue.

As far as precision goes, I guess I was a little bit confused about why maybe it was called apples to oranges. The precision that was presented in the package insert, anyhow, showed a 10 percent. Dr. Rifai brought up the difference that we don't normally see which is higher imprecision and higher values. But it was done using actually 100 points with anackels over 20 days is 80 points. I think that part was a valid study.

Whether 10 percent is adequate or not, I don't know if that was a question that had come up but I think it is. In some tests like cholesterol, for example, where there is a lot of overlap between the disease and the non-diseased group, you need a very

tight C.V.

Three percent has been the goal. When the groups are more separated, certainly we will want to optimize the imprecision to be as small as possible but you are able to tolerate a bit more imprecision when the groups are well separated.

I guess with that I'll stop unless there are other questions.

DR. ROSENBLOOM: I'm not sure that mine was exactly a question but I had some concerns about diabetes.

DR. KROLL: Excuse me. Identify yourself.

DR. ROSENBLOOM: Rosenbloom. I had some concerns about the diabetes group and variations.

DR. BRUNI: We did not break the data out between diabetics and nondiabetics but we have the information.

DR. MAISEL: I just wrote down some notes that I probably can't read but let me try to answer a few things.

DR. KROLL: Could you state your name again?

DR. MAISEL: I'm sorry. Dr. Alan Maisel. We are actually doing this study in diabetes where we're looking at the relative risk of BNP in

predicting cardiac dysfunction in patients with diabetics. I think it's a real exciting area. Myself and Dr. Alan Garber have noted diabetologists are coordinating in the effort.

I did look in our two clinical studies in the ER and the echo studies and there were a lot of diabetics. They just seemed to fit this same pattern. There weren't quite as many Black Americans in those studies but we had about 20 percent.

I did not break down Black hypertensives so I couldn't tell you in particular. Again, when they have heart failure, the levels are way up. When they don't have heart failure in the emergency room, they come in for some other reasons, the level is way down.

I appreciate Dr. Brinker's statements which I wanted to just reply to as well. Does the assay reliably measure the peptide? I think they explained that it reliably does.

Does the level of the peptide reliably tell you what's going on in the heart? It definitely tells you what's going on in the heart. If there are any questions with the age controls or whatever, look in the literature. There's tons of literature.

BNP has been used in Europe. It's been

used reliably, Dr. Packer, as a screen in Edinborough and in several other places in Europe in primary care and for echocardiography. It's been used in Caucasians. It's been used in Blacks. It's been used in women and they find very little differences in the levels at the low end.

I'm not part of the company but I don't think they are trying to invent something new here. I think all that stuff is known. It's being measured. People are using it a lot all over the place.

I think someone from Dr. Packer's own hospital just sent me a seven-page proposal on one who said he believes that LV diastolic dysfunction BNP could be very important. I think it's Dr. Mauer and wants to really use BNP to study that as a titrate treatment. I think that the likelihood of misinformation from this test is very small. As I told you in our --

DR. KROLL: Excuse me. This is Dr. Kroll. You're talking about that and that's an area I think we all brought up several times. One thing I did not see in the packet was propagation of errors or an evaluation or analysis of when there's an error made.

We looked in the data that the FDA submitted that when you brought it up, and this is not

great data to begin with, that if you got the specificity up to 95 percent, the sensitivity is down to 77 percent.

What happens when you take in different types of populations, different types of prevalence groups, and look at wrong assessments coming up based on the data? That's not been available and I think it's one of the things that the committee has brought up several times. There has been concern and there is no analysis and there's no data to support what goes on, at least in the presentation.

DR. MAISEL: Again, I can't say what's in the PMA, what is completely there, because I wasn't totally a part of putting it back together. I was asked to do a number of studies so I could decide for myself whether I would recommend that this be used.

In clinical settings, not just where you are collecting from a lot of people and then trying to put it together, but how do you use it clinically? I tried it clinically and exactly how it's already being used in Europe and it's going to be used here.

I took very sick people coming in emergency rooms where a decision whether it was heart failure or not and it could be a life or death decision. If you get to BNPs under about 55, nobody

with dyspnea had congestive heart failure period.

DR. KROLL: Excuse me again. I would like to remind you we are trying to have each person on the panel have each of their questions from the sponsor answered. I would actually like to go to Dr. Henderson now and see if her question has been answered.

DR. MAISEL: Okay. I'm sorry. I apologize.

DR. HENDERSON: No, my question was answered. However, I just want to have a comment on what you were saying. What I would infer from what you just said in the previous round of comments, it sounds as though you believe that the low values are fairly consistent across all age groups and any subgroups that have been studied. Would you then say that as a negative predicting test that it is very good? Is that your assessment of all of this?

DR. MAISEL: I would say for me seeing patients who come in with dyspnea in the emergency room the best thing you can have in a test is a very strong negative predictive value and a reasonable positive predictive value.

I heard somebody was sort of saying something bad about PSA but I'm glad we have PSA. I

think that this test in a way is sort of like PSA because there's a value under which you are pretty sure that the cause of their dyspnea is not heart failure period. It doesn't matter what group they're in because if anything when you get under 55, you really don't see many people at all, hardly anybody with heart failure that's under 55.

Yes, there is an area somewhere around over 50 and around 100, at least in the ones that we looked at, where you do see some patients that it may have to do with hypertension, may have to do with ethnicity.

Then when you see the really sick people, just like you have really high PSA levels for prostrate cancer, when you see really sick people down in the ER with dyspnea, Milton, I can't comment on the PMA. You may be absolutely right in what you looked at there.

There you're taking a broad sample from a lot of laboratories, you're sending the blood in and you're reporting things on forms, and I was right there doing it right where a life or death situation occurs. The positive predictive value in those cases is very high because when you don't have heart failure, your picograms/mL average 37. When you have

heart failure it's over 1,000.

Now, when you have ROC curves, I agree with what Dr. Packer said about that the age match wasn't perfect here. There are ROC curves all over the literature for BNP and they are all in the high 90s. From what I remember, that's higher than what a PSA ROC curve is. That's higher than a mammogram. It's higher than a cervical --

DR. KROLL: Excuse me again. Let's focus on this thing. Let's go to Dr. Brinker. Do you have any questions that you'd like the sponsor to answer that they haven't answered yet?

DR. BRINKER: No, I don't think so.

DR. KROLL: Dr. Manno.

DR. BRUNI: I've got a comment.

DR. KROLL: Okay.

DR. BRUNI: In response to Dr. Henderson's question, I thought ethnicity might play a role and the difference between women. We tested additional Black women and compared Caucasian to African-American women. If I can have slide -- this information is not in the PMA. It's something I followed up on as a result of some of the differences that we measured.

DR. GUTMAN: You can't actually present it if it's not in the PMA. You can describe it but you

can't present it.

DR. BRUNI: Okay. I'll describe it. We tested 50 African-American women and 106 Caucasian women.

DR. HENDERSON: Were they hypertensive?

DR. BRUNI: None of them were hypertensive. All of these were apparently healthy with normal blood pressure. The mean concentration was 17.6 in the Caucasian population and 18.5 in the African-American population, mean being 12 and 14.

DR. ROSENBLOOM: What were their mean ages?

DR. BRUNI: Mean ages, I did not calculate that. I'm sure it's representative of the same sort of population.

DR. HENDERSON: Thank you.

DR. KROLL: Dr. Manno, did you have any unanswered questions?

DR. MANNO: There was one question or statement or comment that I forgot to include in my other round.

DR. KROLL: Okay. Why don't you ask it.

DR. MANNO: One was there was a statement in the data packs that they tested the pipettes for the volume deliberated. It really didn't make a

difference whether you used 200 microliters or 300 microliters of samples. Therefore, they didn't figure that volume of sample made much difference, but there is no data for other volumes that may have been tested.

It's been my experience that is not the case with pipette calibration or volume to do the test on. When you are going to do it weight per volume, you've got to have pretty narrow --

DR. BRUNI: Can I have slide 92?

DR. MANNO: You can just basically tell me if you would like.

DR. BRUNI: Well, we varied the concentration from 200 microliters to 300 microliters and tested three control volumes 20 times each, the first control being in the normal range and the second one roughly mid-range in the standard curve, and the third one being in the upper part of the standard curve. You can see that the recovery of BNP did not vary with volume. We provided disposable pipette with the product that delivers 240 microliters.

The next slide shows the relative imprecision of a precision pipette coefficient variation being .6 percent and the coefficient variation of the disposable pipette that we provided

with ours is 3 percent. Three percent of 240 is roughly seven microliters. I think we covered that range.

DR. MANNO: Okay. Thank you.

DR. KROLL: Thank you. It's now 3:00 and we have to open this meeting as an open public hearing. If there are any interested persons who wish to address the panel and present information relevant to the agenda, we would like to ask them to come up now.

MS. CALVIN: Anyone from the public can make comments to the panel.

MR. ROBINSON: My name is Gary Robinson and I'm at Igen. I just have a quick question about the use of the test in the emergency room and the prevalence of undiagnosed CHF among patients presenting to the emergency room. Does the sponsor know what the prevalence, not just in the study that they did but across the country but --

DR. KROLL: I'm sorry. It's not appropriate to be asking the sponsor questions. This is a forum to make comments.

MR. ROBINSON: Okay. The comment is that the prevalence was not described. The actual prevalence among the population of the United States

was not described and that raises a question about the predictive value of the test around the country.

DR. KROLL: All right. Thank you. Is there anyone else who would like to come up and make a comment? All right.

Then we can actually proceed with going around with the panel addressing additional comments or questions they would like to hear addressed by the sponsor.

Mr. Reynolds.

MR. REYNOLDS: Just to clarify my comment concerning the precision testing. If you look at the data that was presented, basically you have the performance at Biosite and then you have performance at the four evaluation sites. Biosite tested three samples, a mean of 29 picograms, 584 picograms, and 1,080 picograms.

Now, Dr. Maisel has already indicated that people with real clinical illness you see very high values. Is that correct? Values as much as 1,000. But at the sites where they did the studies, you had means of approximately 25 and 163. You didn't have a sample in that high range so why wasn't a precision study done on a sample in that high range if that is significant from cladical point of view?

DR. BRUNI: We are going to be providing a control set that the consumer can purchase. This control set was going to contain two controls. One is an elevated control but no elevated that it's sky high. It's within the normal range to show that it's not varying. These controls will be used at the laboratory's discretion and according to the respective regulations on the use of quality control.

Also we have a calibration verification control that contains three controls, one at the high end, one in the middle, and one at the low end so they can verify the calibration. These two products have been cleared by 510(k).

DR. KROLL: Thank you. In terms of using our time effectively we need to eventually turn to answering the FDA questions. What I would like to do is ask the rest of the panel members if they have any questions that they asked before that they felt that the sponsor hasn't sufficiently tried to make an effort to answer.

If they have answered it before and we've heard a response, we potentially could come back to that in our conclusion but, in essence, is there something that you think they need to have additional response to?

DR. PACKER: Milton Packer. I just have one question, hopefully a very brief answer. What cutoff would the sponsor propose to distinguish someone with or without heart failure? What cutoff for BNP in someone who is 65 years old and has hypertension and coronary disease? What value? You can answer that separately for a man and for a woman. What would be the proposed cutoff?

DR. BRUNI: Based on the clinical sensitivity and specificity and the rate of change in coming from, say, 40 to 110 nanograms/mL irrespective of age, including the hypertensive group with the apparently healthy group as being "non-CHF patients," I think something in the neighborhood of 80 to 100 nanograms/mL would be appropriate. Also keep in mind that you're going to have a little lower sensitivity in the asymptomatic patient than you are in the patient that is very sick.

DR. PACKER: According to your own data 40 percent of your normal/hypertensive patients have a value greater than that.

DR. BRUNI: That's true. I would say 100 picograms/mL.

DR. KROLL: All right. Dr. Everett.

DR. EVERETT: Yes. I would like for the

sponsor to make some reply to what the FDA suggest and that is that the device is really safe and effective in only one group, males, and perhaps young males.

DR. KROLL: Please let's keep this succinct because at 10 after I want to answer the FDA questions.

DR. MAISEL: I understand. I have a plane also. Little League tomorrow. In the literature BNP has been shown to be effective in all groups. I believe the PMA doesn't have enough Black Americans to tell you absolutely so I guess you're relying on the fact that there's a device which very accurately measures a test.

To answer your question, sir, everything that's on this device has already been used in many hospitals around the country on an FDA approved platform, triage cardiac platform, so I think all those concerns have been addressed and previously approved by the FDA.

Do they have enough Black Americans in that population? To answer your question with their specific device, they probably don't. I think is it likely that we're going to find out once it's out? Yes. Is it likely that it's going to be pretty much the same because of what's happened in the rest of the

world where they have used BNPs in primary care for a long time? I think it's very likely to be very little. There will be some overlap but not much.

In addition, I must emphasize that the population that was studied here represents the population that was being assessed for congestive heart failure as they showed up to the clinics. We didn't target a particular population of men, women, or otherwise.

As people came in, they consented to participate in the study and the study wasn't biased in any fashion there. I think you will see the same sort of trends regardless of sex or race even though there is a six difference in the normal range. This is what I had mentioned in the PMA that had been reported in the literature.

DR. EVERETT: But you understand my point, though.

DR. BRUNI: I understand.

DR. EVERETT: That's not backed up by your data.

DR. BRUNI: I understand your point, yes.

DR. KROLL: Thank you very much. Now I'd like the panel to specifically look at the FDA questions. Let's have a very brief discussion on each

of these FDA questions. We are going to do this by people raising their hand. Can you project the questions?

MS. CALVIN: Yeah, I have to switch lap tops. I'm sorry. Give me a minute.

DR. KROLL: All right. If you look in the back of the FDA handout, the one that says "FDA Presentation," in the back several pages they have them. I'll read the question while we're getting it up. The first question is that using 55 pg/mL as the final cutoff resulting in the following performance parameters: Age-matched healthy controls versus all patients with CHF; sensitivity 83 percent, specificity 95 percent. Age-matched healthy controls versus patients with CHF (Class I and II); sensitivity 77 percent, specificity 94 percent. Is this the appropriate cutoff or should it be raised or lowered?

We're interested in people's comments. Yes, Dr. Packer.

DR. PACKER: Milton Packer. I think that we already have heard what the limitations are of the age-matched analyses so I am not very comfortable that this represents the sensitivity and specificity in the patient population most likely to be tested.

The sponsor has already said that they

think that a cutoff greater than 55 is appropriate in the patient population at risk, although there would be substantial numbers of patients who would be normal or without heart failure who would have values greater than that. I don't think that we have a basis for deciding on an appropriate cutoff.

DR. KROLL: Thank you. Anyone else on the panel have any additional comments to add?

MS. AMMIRATI: Dr. Kroll?

DR. KROLL: Yes.

MS. AMMIRATI: Erika Ammirati, Industry Rep. I'm most concerned here that there seems to be two spheres of information. There's the one sphere that's more technical, perhaps academic about sensitivity and specificity. I certainly understand that.

Then there's another sphere that Dr. Maisel refers to of, you know, these are people that come through the door and it kind of seems to work. I'm frustrated that the two spheres aren't closer together.

I'm sure we're all feeling that to some extent. I don't even know the value of the comment but I want to at least voice it, that if you look at something very specifically, gee, we wish we had

normal people that are 90, knowing that there probably aren't a lot of normal people that are 90, and we would like them perfectly matched in terms of the group.

There is 25 in each of the various decades, and what that's telling us versus, gee, you know, these people came through the door, they were sick, they couldn't breathe, we measured this, and the answer kind of seemed to work. I just wanted to sort of put that out in the open forum.

DR. KROLL: Dr. Packer.

DR. PACKER: If I could address that. If it works, it should be easy to show. You should be able to design a database in the clinical trial that shows that it works. If it's easy to show, it should be easy to prove to others that it works. The limitation here is not with the device. The limitation is with the database.

DR. KONDRATOVICH: May I show you --

DR. KROLL: Use the microphone.

DR. KONDRATOVICH: May I show the calculation of different cutoffs for different groups only normal versus CHF? I would like to pay attention that with increasing of age cutoff must increase.

DR. KROLL: Can we do this quickly? I

think you may just want to make a comment that as age increases, that perhaps the cutoff might be increased. Let me make a comment myself which is something that we used to say when I was in medical school and we had a note taking service, and that is you can't make steak out of hamburger. I think I agree with Dr. Packer that we don't really have the data here.

DR. KONDRATOVICH: For example, for specificity 0.95 for people under 45 years old cutoff 40 and sensitivity 0.78. For people age group 46 to 65 years old, cutoff 50, sensitivity 0.86. Over 66 years old cutoff 75, sensitivity 0.90. Of course, confidence interval is relatively weak.

I again pay attention that this is only normal versus CHF. Hypertensive, of course, will be bigger cutoff. Therefore, cutoff increase with age.

DR. KROLL: Let me ask Dr. Gutman, do you think we have sufficiently answered this question for you?

DR. GUTMAN: Yes. I can point out -- you know, I won't comment on whether there's steak or hamburger here. There are lots of innovative techniques that statisticians can do to help. You can stratify by age. You can throw in equivocal zones. There are things here you can do. You can ask for

extra data. You can change claims. Any of them are up for grabs as you are giving us advice.

DR. PACKER: But statisticians can only analyze data that they have.

DR. KROLL: Thank you. Let's go on to question 2. That question is the study design was a model studying a pre-selected population (healthy controls, hypertensives, and patients with defined CHF). Although results closely approximate sensitivity and specificity reported in the literature, the test was not studied in actual emergency room use. Should this be indicated in the labeling?

DR. BRINKER: May I ask a question?

DR. KROLL: We're trying to answer this question.

DR. BRINKER: It's germane to these questions.

DR. KROLL: Sure.

DR. BRINKER: The question is going to be to the FDA about these questions. All these questions pretty much pertain to labeling issues as opposed to whether this device, in your eyes, meets the criteria for approval. I think if we work -- is that the platform?

The labeling issues are usually something that is very compromisable at your level even to the term of saying that there is no absolute cutoff value. Here is what this study errored as it is shown and here are the ROC curves and more data will be -- you know, they're charged with more data and to revise the labeling.

DR. GUTMAN: I'd be happy to -- I'm an honest man and believe in truth in labeling so I'll just put everything on the table from my perspective. These questions if you read them look as though they are leading towards an approval because the team was probably satisfied with the notion that it meant the least burdensome threshold that Phil Phillips spoke about.

I'm sorry to say that because we don't wish to -- I mean, you're brought here specifically to quality control us, to give us your best advice and so I actually don't wish to influence you. I want you to give me your fair and square honest advice on where to go with this and you have lots of choices. You can not approve it or you can approve it with the requirement for additional studies. You can approve it that those additional studies can occur before we approve the product. They can occur after we approve

the product.

We did give some credence to the literature. I don't know the literature but if there's a strong literature base suggesting that there isn't a racial difference, we probably wouldn't normally ask a sponsor to go into complicated efforts to demonstrate what wasn't there unless in some subgroup analysis there was.

Now, maybe we're not vigilant enough. There have been at least two analytes that I know of where we didn't pick it up in early development. One was frankly CK and the other was PSA where the racial differences came out only after it went into the field so we'll be wrong again I imagine. There will be all kinds of nuances that we don't pick up.

If we thought there was a literature base that didn't suggest a need to look at racial subtypes, we wouldn't probably push the sponsor. Maybe we should. The deal here is these answers you should answer fair and square but the bottom line at the end of the day is to advise us whether they met the threshold to be safe and effective with this data set and with some kind of appropriate labeling or whether they haven't. If you say they have, what should we do further? If you say they haven't, what should they do

further?

DR. BRINKER: I think this question is important because if we can first -- it may be moot to go over all these questions and then end up saying we don't think it's approval because the data set doesn't mandate it. Maybe we should work the other way around and decide whether there's reason on the basis of the data that we have to suggest it could be approved with some labeling and some additionals post-market or whatever. If not, say why and then these questions become moot.

DR. GUTMAN: It's the Chair's prerogative. I have no objection to whatever approach you take. Wherever we go with the submissions, these questions will be important for us to be answered because we will continue to work with the sponsor.

Whether you approve it, don't approve it, approve it with conditions, we will work with the sponsor to try and get the data and the labeling right. I hope you don't leave without at least addressing some of the questions but the order is immaterial to me.

DR. KROLL: I would recommend that we attempt to answer these questions and think of them isolated from how we feel about or think about

approval or disapproval and just answer them in terms of labeling and what do you think should go in there considering the background. Then let's move on to the final recommendations and vote.

If you want to do this by 4:00, we need to do it succinctly. Again I open up the question for No. 2, the issue about whether or not the information that this was not done in an emergency department, whether that's relevant to put into the package labeling.

DR. CLEMENT: I'll say no based on the data. We clearly heard all the data that we're looking at here before us was basically done through cardiac clinics and so forth.

DR. KROLL: I guess what they are asking is --

DR. GUTMAN: I think what we're trying to portray here we were cognizant of the fact it was a non-naturalistic study and we are implying that we are willing to live with that as long as there is cautionary labeling. If the panel wished to take a more extreme view, they could say that this isn't a satisfactory study. That wasn't the answer we expected to hear. That's an honest answer, however.

DR. ROSENBLOOM: Rosenbloom. Doesn't the

QC in the hospital take care of this? I mean, you can't do a bedside test in the hospital without having the laboratory involved handling the QC. I agree with Steve that we don't have to recommend that.

DR. GUTMAN: I think there are two issues mixed up here. One is the issue of performance in the hands of laboratorians versus point-of-care. That's an easy analytical study that can be rectified. If it's not in the submission or if you haven't seen it, one is showing it works as well in the hands of a bunch of untrained people.

That's not what this question is about. This question is about the selection of patients who were actually studied. It wasn't that you took a 1,000 people presenting to an ER, ran the test, and then defined the end point independently. You took selected healthy people and selected hypertensive people and selected congestive heart failure people.

Those are biased samples. We thought there should be some statement about the nature of that bias so people would understand that whatever estimates you got, even if you age-matched successfully, were, in fact, perhaps imprecise or crude or only ballpark.

DR. KROLL: My answer to that is I think

you should include that information so people reading it are aware but you are obviously getting a mixed group here. I think it's difficult for people to say. Any other comments on question 2?

DR. RIFAI: I think it's important to note such a statement. In my mind it's more important that the actual test was done by professional laboratory people so this is the best case scenario you are going to find.

DR. KROLL: Dr. Gutman, have we sufficiently answered the question?

DR. GUTMAN: We don't mind diversity of opinion. That's okay.

DR. KROLL: All right. Let's go to question 3. If we could put that up. I don't know if I want to continue reading all of these. Can everybody read this question? I'll read the last line which is what should be included in the labeling to ensure that users understand the potential variable performance of the assay?

This is really addressed to looking at different groups. I think the answer to this question is irrelevant whether we think this has been answered today to our satisfaction. The question is what would go into the labeling if we lived in a perfect world

and we could get perfect data. What would be want to put in there.

DR. HENDERSON: Does this question refer to the differences in the group studies and assayed or this refers to the test itself in the laboratory?

DR. GUTMAN: It's the sensitivity and specificity calculations change depending on what you're looking at. As our statistician showed you, if you compare, for example, the congestive heart failure I and II, you get different sensitivity and specificity estimates than if you compare the whole group, I think, even when you age-matched.

The issue is should all that data be put in? Should the worse case scenario be put in? Should the best case scenario be put in? Should we average the data? One of the challenges is you can make package inserts longer and longer and more and more comprehensive and then nobody understands what's in them.

DR. BRINKER: I think you should supply the data, all the comparisons, but have a summary point in summarizing. People should be able to understand what the intricacies of the clinical experience has been.

DR. KROLL: All right. Any other comments

on that particular question?

DR. EVERETT: This is James Everett. I think that it should be in there, particularly if the performance varies based on where the test is done, whether it's done in an emergency room or in the laboratory. Actively I do ER work and there are just some things that happen in the emergency room with tests that just don't resemble what happens when the lab does a test.

In the hospital it's a constant battle. Who's doing it right and who's doing it wrong or is it an inherent problem with the test or where it's stored when we get ready to do the test. If there is something that definitely affects the performance, then I think that should be in there.

DR. KROLL: All right. Any other comments? Dr. Gutman, we sufficiently answered question 3 for you. Let's go to question 4. FDA has evaluated the cutoff using age-matched data and ROC curves. Is this the appropriate analysis? I assume were you're talking about the FDA has done this, you're talking about the presentation we received today. Do you have other suggestions on how data should be analyzed and presented? Dr. Packer.

DR. PACKER: I don't think that this is